

Torsten Schlieder

Kollegiat seit: 01.10.1999
Erstbetreuer: Prof. Dr. Heinz Schweppe
Zweitbetreuer: Prof. Dr. Myra Spiliopoulou

Telefon: +49 (30) 838 75 135
E-Mail: schlied@inf.fu-berlin.de
Homepage: www.inf.fu-berlin.de/~schlied



Thema: Similarity Search in XML Data

Abstract:

The objective of my work is to help bridging the gap between the expressiveness of structured queries and the vagueness of information retrieval with respect to XML data.

I introduce the simple query language *approXML*, which supports hierarchical, Boolean-connected query patterns. The interpretation of *approXML* queries is founded on cost-based query transformations. The summary cost of a sequence of transformations measures the similarity between a query and the data. The similarity score is used to *rank* the results. The language allows the renaming of query elements, the insertion of elements into the query, and the deletion of a restricted set of query elements. Each type of transformation has its intuitive semantics: The renaming of a query element *changes* the search context of a query part. The insertion of a query element restricts a query part to a *more specific* context. Finally, the deletion of a query element changes the search space to a *more general* context. The costs of the basic transformations renaming, insertion, and deletion are specified by the a domain expert. All results of an *approXML* query can be computed in polynomial -- typically sublinear -- time with respect to the database size.

Publikationen:

[Schlieder, 2001a]

T. Schlieder. *ApproXML*: Design and implementation of an approximate pattern matching language for XML. Technical Report B 01-

02, Freie Universität Berlin, May 2001.

[Schlieder, 2001b]

T. Schlieder. Similarity search in XML data using cost-based query transformations. In *Proceedings of the Fourth International Workshop on the Web and Databases (WebDB'01)*, Santa Barbara, USA, May 2001.

[Schlieder and Meuss, 2000]

T. Schlieder and H. Meuss. Result ranking for structured queries against XML documents. In *DELOS Workshop on Information Seeking, Searching and Querying in Digital Libraries*, Zurich, Switzerland, December 2000.

[Schlieder and Naumann, 2000]

T. Schlieder and F. Naumann. Approximate tree embedding for querying XML data. In *ACM SIGIR Workshop On XML and Information Retrieval*, Athens, Greece, July 2000.

[Schlieder, 2000]

T. Schlieder. Strukturelle Ähnlichkeitssuche in XML-Dokumenten. In *Tagungsband des 12. Workshop "Grundlagen von Datenbanken"*, pages 91-95, Plön, Germany, April 2000.